Effects of Display Type on Performance in Virtual Environments

Donald R. Lampton, John P. Gildea,
Daniel P. McDonald, and Eugenia M. Kolasinski
U.S. Army Research Institute

October 1996

DTIC QUALITY INSPECTED 2

19970306 012

**United States Army Research Institute
for the Behavioral and Social Sciences**

U.S. ARMY RESEARCH INSTITUTE

FOR THE BEHAVIORAL AND SOCIAL SCIENCES

**A Field Operating Agency Under the Jurisdiction
of the Deputy Chief of Staff for Personnel**

EDGAR M. JOHNSON
Director

Technical review by

Cathie E. Alderks
Bob Witmer

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1996, October	2. REPORT TYPE Interim	3. DATES COVERED (from. . . to) September 1994-May 1996

4. TITLE AND SUBTITLE Effects of Display Type on Performance in Virtual Environments	5a. CONTRACT OR GRANT NUMBER
	5b. PROGRAM ELEMENT NUMBER 0602785A

6. AUTHOR(S) Donald R. Lampton, John P. Gildea, Daniel P. McDonald, and Eugenia M. Kolasinski	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 2111
	5e. WORK UNIT NUMBER H01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-IF 5001 Eisenhower Avenue Alexandria, VA 22333-5600	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Technical Report 1049

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT *(Maximum 200 words)*:

   This research was conducted as part of a program to determine interface requirements for enabling dismounted soldiers to train in Virtual Environments (VEs). We compared different VE display devices in terms of their effects on task performance, skill acquisition, and side effects. Forty-eight college students completed a series of visual and psychomotor tasks, a subset of the Virtual Environment Performance Assessment Battery (VEPAB), using either a Head-mounted Display (HMD), a head-tracked boom-mounted display, or a standard computer monitor. Performance on vision tasks was sensitive to differences in display devices and to individual differences. Visual acuity scores were ordered according to estimates of the resolution of the displays, but were worse than what would be predicted from the resolution estimates. In comparison to real-world performance, distance and height estimation in the VEs varied greatly across participants, especially with the HMD. Motor tasks had high reliability, demonstrated small but significant practice effects, and were correlated with participants' use of computers and video games. Unexpectedly, even the standard monitor group showed a significant increase in simulator sickness scores. The VEPAB tasks should prove useful in the future when design tradeoffs must be made in the process of developing training system prototypes.

15. SUBJECT TERMS

Dismounted infantry	Training	Virtual reality	Human performance
Virtual Environments	Simulator sickness		

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	53	

i

Effects of Display Type on Performance in Virtual Environments

Donald R. Lampton, John P. Gildea,
Daniel P. McDonald, and Eugenia M. Kolasinski
U.S. Army Research Institute

**Simulator Systems Research Unit
Stephen L. Goldberg, Chief**

October 1996
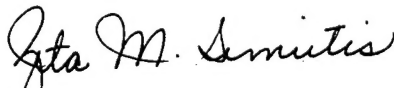
FOREWORD

The U.S. Army has made a substantial commitment to Distributed Interactive Simulation (DIS) and the electronic battlefield for training, concept development, and test and evaluation. The current DIS training system, Simulation Networking (SIMNET), and the next generation system, the Close Combat Tactical Trainer (CCTT), provide effective training for soldiers fighting from vehicles, but are unable to do the same for individual dismounted soldiers. Virtual Environment (VE) technology has the potential to provide Individual Combat Simulations for the electronic battlefield. However, several research challenges must be overcome before VE technology can be used for practical training applications. These challenges include: providing all trainees with the necessary prerequisite skills for operating in VEs; identifying and quantifying the effects of VE system characteristics that influence skill acquisition and transfer; and making sure that unwanted side effects and aftereffects which might result from immersion in VEs do not pose unacceptable risks.

This report describes research conducted to examine performance in immersive virtual environments as a function of different visual display devices. The results of the research provide information about the characteristics needed for effective learning and performance in virtual environments.

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) Simulator Systems Research Unit conducts research to improve the effectiveness of training simulators and simulations. The work described is a part of ARI Research Task 2111, VIRTUE - Virtual Environments for Combat Training and Mission Rehearsal.

ZITA M. SIMUTIS
Deputy Director
(Science & Technology)

EDGAR M. JOHNSON
Director

v

EFFECTS OF DISPLAY TYPE ON PERFORMANCE IN VIRTUAL ENVIRONMENTS

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army has made a substantial commitment to the use of virtual environment (VE) technology to create virtual battlefields for combat training and mission rehearsal, development of military doctrine, and evaluation of weapon system concepts prior to acquisition decisions. All of these functions would be improved by a better representation of dismounted soldiers on the virtual battlefield. The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has established a research program to determine the characteristics of the VE technologies needed to allow dismounted soldiers to fight on virtual battlefields. As part of that program, the objective of this experiment was to compare different VE display devices in terms of their effects on task performance, skill acquisition, and side effects.

Procedure:

Forty-eight college students completed a series of visual and psychomotor tasks using either a Head-mounted Display (HMD), a Binocular Omni-Orientation Monitor (BOOM), or a standard computer monitor. The initial level of performance, practice effects, the effects of different visual displays on task performance, and the incidence and severity of side effects and aftereffects resulting from VE immersion were measured.

Findings:

Performance on vision tasks was sensitive to differences in display devices and to individual differences. Visual acuity scores were ordered according to estimates of resolution based on the horizontal pixel density of the displays, but were worse than what would be predicted from the resolution estimates. In comparison to real-world performance, distance and height estimation in the VEs varied greatly across participants, especially with the HMD.

Motor tasks had high reliability and demonstrated small but significant practice effects. Unlike the vision task scores, most of the motor tasks were correlated with participants' use of computers and video games, with higher use associated with better performance. Participants' performance on a search task was faster in the monitor condition.

Unexpectedly, even the monitor group showed a significant increase in simulator sickness scores. We believe that the Virtual Environment Performance Assessment Battery (VEPAB) locomotion tasks induce vection, the perception of self-motion, and that the discrepancy between the visual system, which signals movement on the part of the participant, and other

sensory systems which do not, result in simulator sickness. No changes in postural stability were found.

Utilization of Findings:

The findings have several implications for future research and practical applications involving training with VEs. First, empirical measures of VE visual acuity are needed in addition to physical measures of the display device itself. For each device, acuity was worse than would be predicted from the resolution of the device, and acuity varied across individuals using the same device. Second, effects of display device characteristics are related to task requirements. Performance on simple locomotion tasks did not vary as a function of display characteristics whereas size and distance estimation varied greatly across devices. Finally, the use of standard monitors to display VEs does not completely alleviate simulator sickness concerns.

The VEPAB tasks have proven to be useful for assessing the effects of interface characteristics on human performance in virtual environments. VEPAB can be used to provide a general orientation for interacting in VEs and to determine both entry level performance and skill acquisition of users. In addition, VEPAB allows comparison of task performance, side effects and aftereffects, and subjective reactions across different VE systems. By providing benchmarks of human performance, VEPAB can promote continuity in training research involving different technologies, separate research facilities, and dissimilar subject populations. The VEPAB tasks or similar tasks should prove useful in the future when design tradeoffs must be made in the process of developing training system prototypes.

EFFECTS OF DISPLAY TYPE ON PERFORMANCE IN VIRTUAL ENVIRONMENTS

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

EFFECTS OF DISPLAY TYPE ON PERFORMANCE IN VIRTUAL ENVIRONMENTS

Introduction

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has established a research program to determine the characteristics of the Virtual Environment (VE) technologies needed to provide an interface for dismounted soldiers to train with Distributed Interactive Simulations such as the Close Combat Tactical Trainer (CCTT). A goal of the program is to determine those technologies that produce cost-effective transfer of training from VE practice to real-world performance, and strategies for the effective use of those technologies.

To support this research, ARI developed a set of simple generic tasks, the Virtual Environment Performance Assessment Battery (VEPAB), to measure human performance in VEs (Lampton, Knerr, et al., 1994). The VEPAB was designed to provide a general orientation for interacting in VEs, to determine both entry level performance and skill acquisition of VE system users, and to allow comparison of task performance, side effects and aftereffects, and subjective reactions across different VE systems.

Four experiments were conducted using VEPAB tasks to investigate the effects of interface device and practice on performance. This report describes the last of those experiments, which examined the differences in performance resulting from the use of different visual display devices. This topic was of particular interest because there are a number of options available at a wide range of prices. Computer monitors provide high resolution full color views at a cost of less than $2000; however, they provide fishtank rather than immersive views. LCD-based Head-mounted Displays (HMDs) cost about $8000. They provide full color, immersive views, but at low resolution. Head-tracked displays can use smaller counter balanced CRT-based displays to provide high resolution, but are expensive (around $80,000), cumbersome, and some versions do not provide full color.

Military Training Requirement

The U.S. Army has made a substantial commitment to the use of networked simulators to create virtual battlefields for combat training (Alluisi, 1991; Sterling, 1993). VE training systems in use, such as Simulation Networking (SIMNET), and under development, such as the CCTT, provide training for soldiers fighting from within vehicles such as tanks and infantry fighting vehicles. In these training systems, crew members operate inside physical mockups of combat vehicles with video monitors providing their views of the simulated outside world (the virtual battlefield) as seen through vision blocks of armored vehicles, cockpits of aircraft, and the electro-optical sights of weapon systems. Trainees crouch, sit, or recline in the simulators in positions similar to those they would assume in actual combat vehicles. Hidden speakers convey appropriate battlefield sounds. Thus, the ways in which the crew members view and hear the battlefield, control vehicle movement, and employ weapon systems correspond closely

1

to those of actual combat vehicles. In marked contrast, the representation of dismounted infantry is limited to a unit leader, seated at a workstation, controlling a group of icons that represent dismounted soldiers. This approach may train mounted soldiers to fight with and against dismounted soldiers, but it is inadequate for training the dismounted soldiers themselves.

Gorman (1990) proposed that immersive VE technologies could provide an interface to enable dismounted soldiers to train on virtual battlefields. With this approach, an HMD would present a view of a computer-generated, three-dimensional environment relative to an eye point within the environment. The human user could control the direction and movement of the eye point and interact with simulated objects within the VE. In this manner the user is "immersed" in the VE. This interface would allow trainees to move, shoot, and communicate in VEs. Simulation of these three basic soldier functions would enable training on complex scenarios requiring real-time tactical decision making.

In addition to training applications, virtual battlefields can support several other functions important to the U.S. military (Alluisi, 1991). For example, virtual battlefields can be used to develop and evaluate military doctrine and to evaluate weapon system concepts prior to acquisition decisions. All of these functions would be improved by a better representation of dismounted infantry.

Virtual Environment Technologies

In a review of immersive VE technologies available in 1992, Levison and Pew (1993) concluded that high costs and performance limitations, such as the limited resolution of HMDs, precluded the immediate, widespread application of those technologies for training dismounted soldiers at that time. However, they also noted that rapid advances were being made in the areas of visual and auditory displays. Pimentel and Teixeira (1993) have pointed out that absolute realism may not be necessary to create a sense of immersion; the created world need only be real enough for the user to suspend disbelief for a period of time.

We believe that our research with the admittedly limited immersive VE technology available today can guide the timely and cost-effective development of VE training systems as technology improves. Our research with low-end VE systems is especially relevant for military training in that networking and cost requirements will constrain the selection of VE systems. For example, in networked training simulations there will be limits on the kinds and amount of information that can be passed across the network. In addition, the requirement to network many trainees simultaneously will place an emphasis on using inexpensive equipment. A long term goal of our research is to determine if immersive VE training can enable the learning and practice of skills that transfer to field training and can be expected to transfer to actual combat.

ARI has established a research program to determine the characteristics of the VE technologies needed to provide an interface for dismounted soldiers to train with virtual

battlefields such as SIMNET and CCTT. As part of this effort, ARI contracted with the University of Central Florida's Institute for Simulation and Training (IST) to develop a laboratory for the conduct of psychological research addressing human performance in VEs. IST provides computer science expertise in the development and operation of the laboratory, acquiring off-the-shelf system components such as hardware and software for the laboratory, or developing them when necessary. Development of the test bed is described in greater detail by Moshell, Blau, Knerr, Lampton, and Bliss (1993).

The Virtual Environment Performance Assessment Battery

Several research challenges must be overcome before VE technology can be used for practical training applications. These challenges include: (a) providing all trainees with the necessary prerequisite skills for operating in VEs; (b) identifying and quantifying the effects of VE system characteristics that influence skill acquisition and transfer; and (c) making sure that unwanted side effects and aftereffects which might result from immersion in VEs do not pose unacceptable risks. VEPAB provides a set of standard materials and procedures to investigate these issues. The materials are simple VEs in which movement, tracking, object manipulation, and reaction time tasks can be performed. The procedures are the instructions, presented to VE users, which describe the tasks to be performed and the measures of task performance. The rationale for the development of VEPAB is described in Lampton, Knerr, et al. (1995).

The VEPAB consists of a set of simple generic tasks that represent components of more complex activities. Simple tasks, in contrast to detailed training scenarios, are easier to develop, provide a general context applicable to other areas of training research, allow isolation of critical variables, and facilitate measurement of performance. We developed tasks for each of five categories: vision, locomotion, tracking, object manipulation, and reaction time. These tasks are components of the soldier tasks in which we are interested. They are components of many nonmilitary activities as well.

Although the VEPAB tasks are clearly related to soldier functions, the tasks were designed so that previous military training is not required for successful task performance. Therefore civilians, such as college students, can be used as research participants. In order to implement the tasks quickly and inexpensively, we simplified several aspects of the tasks and the interface with them. Only visual cues were provided. No auditory or haptic devices were used. The participant's body was not visually represented in the VE. Manipulation tasks were performed using the same control devices used for the other tasks. The use of more sophisticated display and control devices (such as three-dimensional sound, instrumented gloves, and instrumented treadmills), and visual representation of the human figure are important areas for future research; however, their investigation was beyond the scope of our initial efforts. Table 1 provides a summary of the subset of the VEPAB tasks used in this experiment.

3

Table 1

Virtual Environment Performance Assessment Battery Task Descriptions by Category

TASK CATEGORY	TASK NAME	TASK DESCRIPTION
Vision	Acuity	Read letters in a Snellen eye chart
	Color	Detect colors in Ishihara plates
	Object Recognition	Identify an object (a human figure) at the end of a 40-ft hallway
	Size Estimation	Estimate the height of a human figure at the end of a 40-ft hallway
	Distance Estimation	Indicate when the image of a human figure, moving toward the viewer from an initial distance of 40 feet, is 30, 20, 10, 5, and 2.5 ft away
	Search	Detect a target moving about the walls, floor, or ceiling of a room
Locomotion (walking)	Straightaway	Move down a straight corridor to a circle on the floor, turn around, and return to the starting point
	Back up	Move down a straight corridor to a circle on the floor, then move backwards to the starting point
	Turns	Move through a corridor formed by 10 alternating left and right 90 degree turns
	Figure-8	Move around a figure-8 shaped corridor
	Doorways	Move through a series of rooms connected by doorways that are offset so that a curved course must be followed
Tracking	3-D Moving Target	Use a control device to move a cursor onto a target which moves in three dimensions
Manipulation	Bins	"Grasp" a ball located in a vertical rack of open containers (bins), pull it out of the original bin, and push it into a target bin
Reaction Time	Choice	Indicate in which of four boxes an "X" has appeared

Vision Tasks

The vision tasks include Acuity, Color, and Search. Other vision tasks address the recognition of a familiar object, a human figure, and the estimation of the size of, and distance to, the object. These visual skills are important to the performance of many real-world tasks. The vision tasks require participants to report aloud what they perceive.

Acuity. A Snellen eye chart, a virtual reproduction of the familiar letter chart found in every doctor's office, was created by modeling each individual letter. The chart is presented at eye level at the end of a 20-ft corridor. From twenty feet away the correct identification of only the top letter of the chart corresponds to an acuity of 20/200. We anticipated that the resolution of some VE visual displays would result in acuity worse than 20/200. To enable the measurement of acuity worse than 20/200, the VEPAB Acuity task allows the eye point to be moved toward the chart in one-foot increments. (The participant's real-world acuity is measured before this VE test is administered.)

Boff and Lincoln (1988), Olzak and Thomas (1986), and Geldard (1972) provided overviews of the concept and measurement of acuity relevant to the VEPAB Acuity task. The ability to perceive spatial detail clearly is termed visual acuity. To assess acuity, high-contrast patterns of various sizes are presented to an observer at fixed distances. The smallest pattern to be detected or identified is taken as the threshold value. Acuity is expressed in a number of ways. Decimal acuity is the reciprocal of the threshold value in minutes of arc, and is expressed in terms of a decimal. Normal acuity expressed in decimal notation is 1. Acuity can also be expressed as a Snellen fraction. The denominator of the fraction is the distance at which the optotype (the Snellen letter stroke) subtends 1 minute of arc. The numerator of the fraction is the actual distance from the observer to the optotype. The most commonly used method for assessing acuity involves reading a Snellen eye-chart placed at a distance of 20 feet. This obtained threshold is considered a person's far point acuity. Expressed as a Snellen fraction, normal far point acuity is 20/20. Near point acuity is usually measured at 40 cm. Acuity measures obtained at these distances may differ. For example, older persons may be unable to accommodate (focus sharply) on close objects.

Obviously, a critical determinant of the resolution of a VE display is the number of pixels per degree of the field of view (FOV) (Padmos & Milders, 1992). Boff and Lincoln (1988) listed accommodation (the degree to which an individual can properly focus on the display), luminance, and contrast as important display factors that affect acuity and pointed out that it is difficult to predict the interaction of factors that affect acuity.

Color. Standard Ishihara color vision test plates were digitized to produce VE tests for mild (red-green) and severe (blue-yellow) color vision deficiencies. Three circles, made up of colored dots, appear on each plate. Dot patterns within the circles form numerals. Participants read the numerals aloud. Scores can vary from zero to three correct for each plate. The

participant's real-world color vision is measured before this VE test is administered. Thus, this task is a gross measure of the capability of the VE system to present appropriate colors.

Object recognition, size estimation, and distance estimation. The visual scene for the distance task was modeled to represent a corridor inside a modern office building. The floor had a checkerboard pattern made up of alternating light and dark one-foot squares, the walls were 12 ft high with narrow vertical stripes every 5 ft, and the ceiling had horizontal light panels every 10 ft. At the beginning of the task, a digitized picture of a human figure (a soldier with a rifle) is shown against a homogeneous gray background at the end of a 40-ft corridor. The participant is asked to identify the object and estimate its height. The participant is told the correct height, 6 ft, and asked to estimate the distance to the figure. The experimenter then tells the participant that the figure is 40 ft away and that the figure will begin to move forward. The participant calls out when the figure appears to be 30, 20, 10, 5, and 2.5 ft "arms length" away.

Search. The participant's viewpoint is at eye-level in the center of a room. A red ball appears near the floor, ceiling, or walls, and moves slowly around the room. Participants search for the target by making head movements and turning their bodies, then call out when the target is detected. For VE systems without head tracking, a manual control device, such as a joystick, can be used to slew the FOV. One practice trial and ten performance trials are conducted. The Search task provides a transition from the vision tasks to motor tasks in that the Search task is presented more than once and a more active role is required of the participant.

Locomotion Tasks (Walking)

Many VE applications will require participants to move at realistic rates while simultaneously attending to other tasks; the cognitive load imposed by walking in VEs should not be significantly greater than real-world walking. We designed a series of progressively more complex locomotion tasks to systematically train participants to move at a reasonable speed while avoiding collisions with obstacles such as walls and door frames. In addition to providing an efficient way to train VE locomotion, the tasks provide objective performance measures, and indirectly provide diagnostics of problem areas.

The locomotion tasks require the participant to "walk" through the VE by using an input control device to direct the speed and direction of the participant's simulated body. We refer to this mode of virtual locomotion as "walking," inasmuch as the height and rate of movement of the viewpoint are constrained to match that of walking, and the contexts (hallways, doorways, rooms) are those in which walking is normally conducted. The dimensions of the virtual body represent the 50th percentile male; the height is 68 in. And elbow-to-elbow breadth is 16.5 in. (McCormick & Sanders, 1976). Eye level is set during pretesting activities, and corresponds to the participant's actual eye height in inches. There is no visible representation of the body. The body can move forward or backward, laterally, or rotate. Software settings, chosen to

6

represent normal walking parameters, control the maximum speed and the rates of acceleration and rotation. The body interacts with the VE through collisions with walls and door frames. Because collisions almost always stop forward progress, movement at a reasonable rate requires emphasis on both speed and accuracy.

Straightaway. The first locomotion task (Figure 1) requires the participant to move down a straight corridor to a target location indicated by a circle on the floor, turn 180 degrees, and return to the starting point which is also indicated by a circle on the floor. The walls at opposite ends of the VE corridor have color and form cues to help the participant orient.

Backup. This task is conducted in the same VE as the Straightaway task (Figure1). The participant moves down a straight corridor to the target location, then moves backwards to the starting point without turning around. Both the Straightaway and Backup tasks can be conducted as discrete trials or as continuous movement. For continuous movement, the participant is instructed to move from circle to circle, following the specified instructions until told to stop. In this experiment the tasks were conducted as continuous movement.
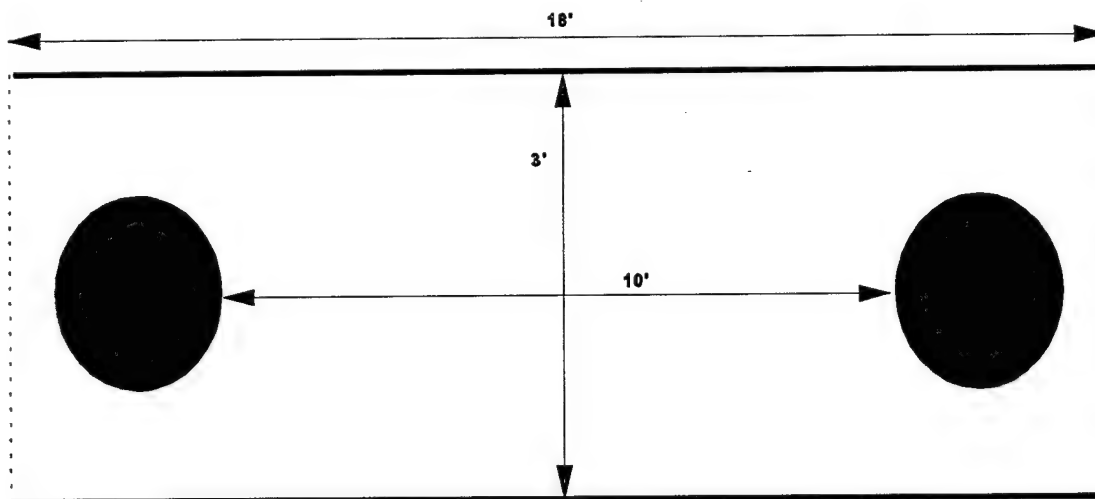
Figure 1. Overhead view of the Straightaway and Backup tasks.

Turns. The task (Figure 2) consists of a continuous narrow corridor (3 ft wide) formed by straightaways joining alternating 90 degree turns to the left and right for a total of ten turns. The lengths of the straightaways vary, with two 20-ft segments alternating with two 10-ft segments.
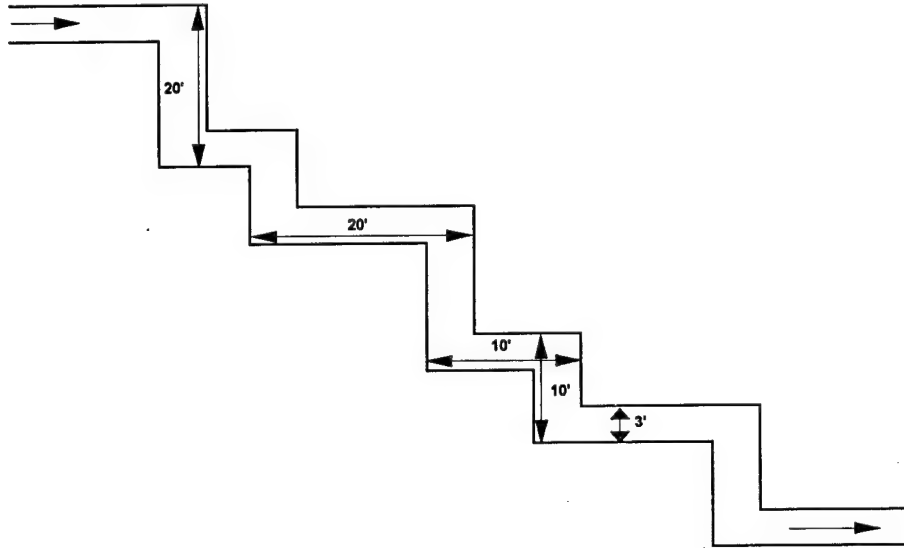
Figure 2. Overhead view of the Turns task.

Figure-8. Two adjoining oval corridors form a figure-8 course (Figure 3). A small diameter oval, rounded on each end with straightaway in the middle, is connected to an oval of larger diameter to form a closed loop. Walking through the course without colliding with the walls requires gradual turns to the left and right. The intersection where the figure-8 course crosses over itself presents a complex visual scene. Figure 3 presents an overhead perspective with arrows added to illustrate the path the participant is to follow. No directional markers are displayed in the actual VE.
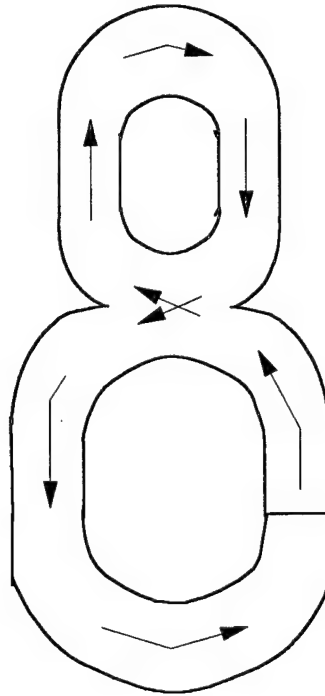
Figure 3. Overhead view of the Figure-8 task.

Doorways. The Doorways task (Figure 4) represents a VE "road test" of the kind and difficulty of walking performance that might be required in a VE training application. The course is formed by a series of 10 rooms connected by 7 X 3 ft doorways. The positions of the doors in the walls vary so that a series of non-90 degree turns must be made to navigate the course efficiently.
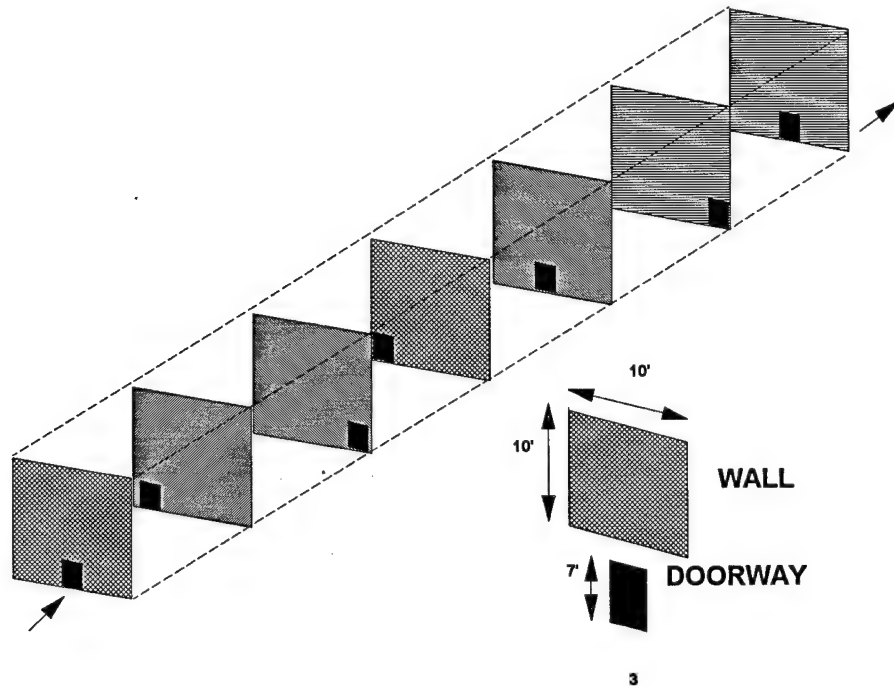
Figure 4. Perspective view of the Doorways task.

Tracking Task. This task (Figure 5) measures skill in controlling the position of a cursor relative to a moving target. The position of the cursor is controlled by the joystick in this pursuit tracking task. The target, a ball .7 ft (8.4 in.) in diameter, appears in a room. It moves in a straight line at about 1.4 ft. per second and changes direction once during the trial. The ball moves in three-dimensional space, that is, up or down, side to side, or toward or away from the participant's VE eye point. The target changes color when the cursor is within the target radius. The target disappears, ending the trial, upon reaching a wall. Random numbers determine the initial position, direction, and change of direction of the moving target. The participant was instructed to maneuver the cursor onto the ball as quickly as possible and keep it tracking on the ball throughout the trial.
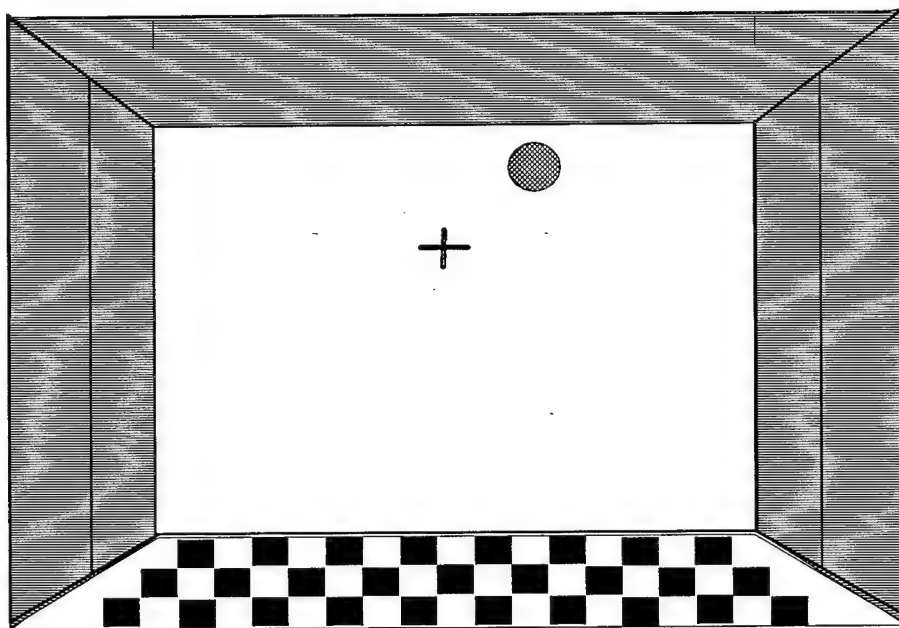
Figure 5. Participant's view of the Tracking task.

Bins. In this object manipulation task, the participant uses a control device to move a cursor which interacts with objects in the VE. The position of the cursor in the VE is shown as a 3-D cross. When the cursor is in contact with the manipulable object, the color of that object changes. Pushing a button on the control device "grasps" the object; a successful grasp is indicated by an additional change in the object's color. A grasped object will move with the cursor.

The participant faces a 3 by 3 stack of open-ended, box-like compartments (Bins). At the beginning of each trial a ball appears in one of the bins, an X appears in another. The cursor is used to grasp and drag the ball out of the bin and into the bin marked with an X. A 45-second time limit is enforced for each trial. Figure 6 depicts the beginning of a trial. The ball is in the left column of the middle row of bins, the cursor is slightly in front of the center bin, and an X marks the target bin, the middle bin in the top row.
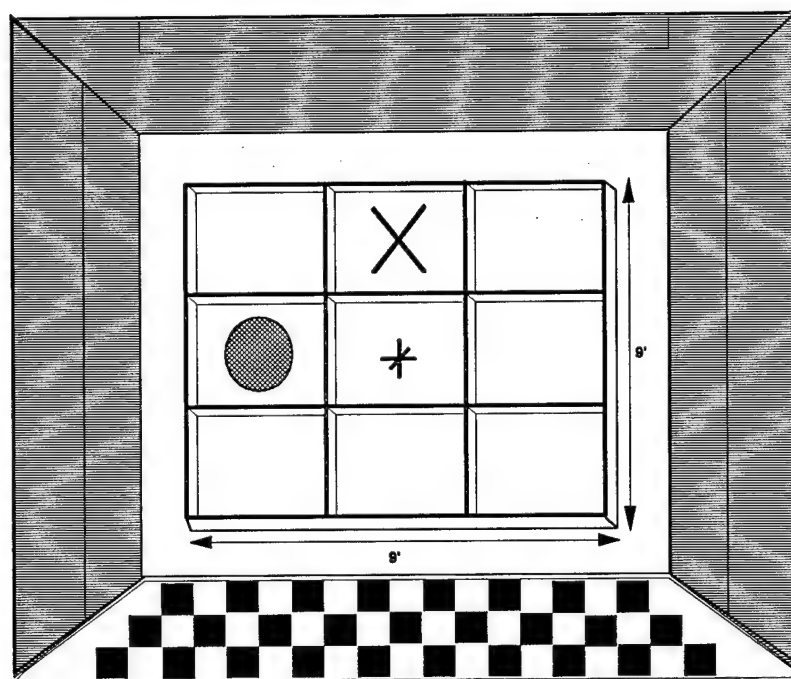
Figure 6. Participant's view of the Bins task.

Reaction Time. The Choice Reaction Time (RT) task (Figure 7) was developed to measure response time to VE events and to provide an indication of the time lag of the VE system in presenting and recording events. For this task the participant pushes the control device in the appropriate direction to indicate in which of four panels an X appears.
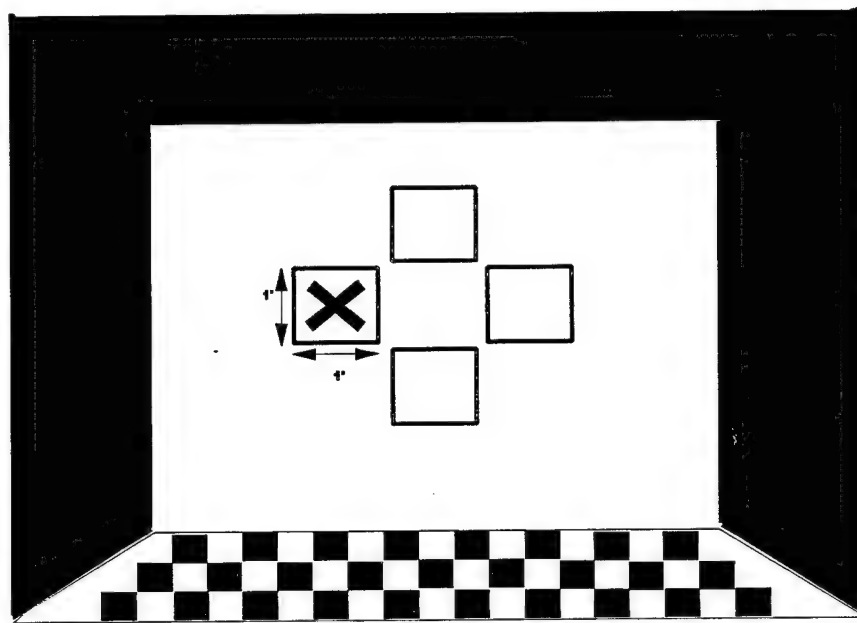
Figure 7. Participant's view of the Choice Reaction Time task.

Research Objective

The object of this experiment was to compare different VE display devices in terms of their effects on task performance, skill acquisition, and side effects. The three commercially available display devices (described in detail in the apparatus subsection of the Method section) represent some of the tradeoffs, such as cost versus resolution or width of FOV, that will have to be considered in the design of immersive VE training systems.

We anticipated that the average visual acuity scores would differ significantly among the three devices as a function of the different resolutions of the devices. A failure to find differences across devices would indicate problems with the VEPAB materials and/or procedures for measuring acuity. For the Search task, we expected better performance with the two devices with head tracking than the device without (a computer monitor). For the other VEPAB vision tasks and the motor tasks, differences across devices were more difficult to predict because of the possible interplay of FOV, resolution, and head tracking.

For all of the tasks, we were interested in individual differences and in correlations among tasks. Of special interest were the correlations among VE and real-world acuity and VE acuity

13

scores and the other tasks scores. These measures would indicate the extent to which VE acuity is related to real-world acuity, and the relation of VE acuity (how well the participant can see in the VE) and motor task performance and side effects such as simulator sickness.

Method

Participants

Forty-eight college students from the University of Central Florida participated in the experiment. They were either paid or received course credit for participation. The age of the participants varied from 17 to 41 years with a mean of 21. Twenty-four were male, and twenty-four were female. An equal number of males and females were assigned to each group. Estimated average weekly computer use varied from 0 to 50 hours with a mean of 8. Participants' estimates of their average weekly video game use varied from 0 to 17.5 hours with a mean of 2.27.

Apparatus

BOOM. The Fakespace Lab^{tm} two-color BOOM2C is a high resolution stereoscopic CRT display. The display contains approximately 1280 x 492 pixels per eye, in pseudo-color mode. Pseudo-color mode produces images that contain mixtures of the primary colors red and green. The BOOM2C display FOV is 140 degrees horizontally and 90 degrees vertically. The display is attached to a mechanical arm that counterbalances the weight of the display. The viewpoint is oriented in three dimensions by six position sensors located at joints in the mechanical arm. As it was designed, the direction of view for the BOOM is controlled somewhat like that of the periscope of a submarine in that the participant peers into the display while grasping control handles on the sides of the display. Each handle has a control button with which the user can interact with the VE. Because the focus of this experiment was on the visual display, we modified the interface with the BOOM so that the participant could use the same control device as that of the other groups, a standard joystick. In order that the participant's hands be free in the BOOM condition, the participant wore a helmet, the liner of a firefighter's helmet, which was strapped to the BOOM. With this arrangement the participant could control the direction of regard by head movement. This allowed a participant to freely survey the VE via head movement while using a standard joystick to perform movement, object manipulation, and reaction time tasks.

HMD. A Virtual Research Flight Helmet was used for the HMD condition. The HMD has an 83-degree horizontal and 41-degree vertical FOV (50-degree horizontal by 41-degree vertical for each eye). The Flight Helmet displays 234 lines of 238 pixels to each eye and uses LCD technology with LEEP optics. A Polhemus IsoTrack provided head tracking of yaw, pitch, and roll.

Monitor. A 20" Hitachi CM2187ME computer monitor with 1024 x 1248 pixels was used for the monitor condition. The monitor did not display stereographic images. Although we made no attempt to fix the participants' point of view relative to the monitor, the normal viewing distance was approximately 24", resulting in a 34.71 degree X 26.94 degree (horizontal X vertical) FOV.

For all three display devices the VE images were generated by a Silicon Graphics Reality Engine. Movement, object manipulation, and reaction time tasks were controlled by a six-degree-of-freedom Gravis joystick.

Procedure

Pretest Activities. The participant read a description of the experimental tasks, possibilities of discomfort, and the schedule of events. The experimenter answered any questions. The participant then read and signed the informed consent. Next, the biographical data form was administered. Then the experimenter used a Keystone Tele-binocular device to administer a set of standard vision tests. These tests measured the participant's acuity, color vision, and stereopsis. Next, the Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Lilienthal, Berbaum, & Hettinger, 1992) and the Immersive Tendencies Questionnaire (Witmer & Singer, 1994) were administered. These two questionnaires were used to establish individual baselines before being immersed in the VE.

The participant completed a test of Postural Stability. For this test, the participant was fitted with a HMD equipped with a Polhemus Fastrack, and instructed to stand with feet heel to toe and arms crossed with hands resting on opposite shoulders. The participant was instructed to remain as stable as possible for thirty seconds. This test provided a baseline of the participant's postural stability prior to being immersed in VE. No visual display was presented during the posture test; the HMD was used only so that the head position tracker could record head movement. Next, the participant was measured for eye height and that measurement was used to set the height of the participant's viewpoint in the VE. The participant was then given operating instructions specific to the display device in which they were to be tested. The pretest activities took approximately 30 minutes.

Test Activities. Participants in each of the conditions were seated in a swivel-base office chair. In the BOOM condition, participants were secured to the BOOM using a set of straps that attached the BOOM to a helmet that the participant was wearing. If the HMD was used, convergence was checked in order to assure the computer-generated images were correctly aligned. The VEPAB tasks were administered in the order listed in Table 1. The Straightaway and Backup tasks were presented only to allow the participants to become familiar with the control device, no performance data were collected. The VEPAB test activities took approximately one hour. Participants were instructed that they could take a break at any time during the session.

Post VE Immersion Activities. After completion of the tasks, the participant was immediately put into the HMD and completed a post-immersion Postural Stability Test. This test was identical to the original Postural Stability Test. Next, the participant filled out a series of questionnaires: the Simulator Sickness Questionnaire, Presence Questionnaire, and Motion Sickness History Questionnaire. Upon completion of these questionnaires the participant was debriefed and any questions were answered. Administrative paperwork was filled out and, if not exhibiting any signs of simulator sickness, the participant was escorted to the exit. Post-test activities took approximately thirty minutes. (If any participant had exhibited symptoms of simulator sickness at the end of these activities, they would have been encouraged to remain until the symptoms abated).

Results

The results for the vision tasks are presented, followed by the motor tasks including the Search task. Reliability analyses, descriptive statistics, and comparison of means across devices are presented for most tasks. In addition, correlations among several of the tasks are listed. Complete data sets were not available for every participant because of error during data collection.

Vision tasks

Acuity. Each of the seven lines of letters on the VE Snellen eye chart was treated as a separate measure of acuity. The farthest distance at which all of the letters of a line could be read by the participant was converted to the minimum angle that the participant could distinguish. The inverse of that angle provided a standard measure of acuity. That is, our metric of acuity is the inverse of the minimum discriminable angle in arc minutes.

Table 2 lists the item-total correlation for the acuity scores for each line of the chart. The corrected item-total correlation is the Pearson correlation coefficient between the score on the individual item (line of the chart) and the sum of scores on the remaining items. Cronbach's alpha, based on the average covariance among the 7 items, was .94. Norusis (1995) stated that a Cronbach's alpha value over .9 is large, indicating that the scale is very reliable.

Table 2

Item-Total Correlations for the Acuity Measure for Each Eye Chart Line

Snellen Chart Line Number	Item-Total Correlation
1	.60
2	.78
3	.88
4	.94
5	.92
6	.92
7	.90

Note: number of cases=43

Figure 8 presents the mean acuity for each of the display devices as indicated by each of the 7 lines on the VE Snellen eye chart. For several of the lines there were ceiling or basement effects. If the participant can read the line at 20 ft, the maximum presentation distance, then the acuity is as good as, or better than, the angle associated with that line. For example, all participants in the Monitor group could read the first and second lines at 20 ft, so the acuity is at least as good as .2 (20/100). In Figure 8, for line 1 the plotting symbols for BOOM and Monitor overlap, reflecting the ceiling effect in that the best acuity that can be measured for line 1 at 20 ft is .1 (20/200). For some participants, a basement effect was obtained with lines 5, 6, and 7. That is, if the participant could not read the line at 1 ft then the acuity is worse than the acuity associated with that line at 1 ft, but we do not know how much worse. Four of the participants, all in the HMD group, could not read line 7 at even 1 ft, the closest presentation distance. Therefore, in the figure, the line 7 mean score for the HMD group is based on only those participants who could read line 7 at a distance of 1 ft or greater. Only lines 3 and 4 did not have any ceiling or basement effects. Because line 4 had a higher Item-Total Correlation, we used the line 4 scores, rather than line 3, for additional analyses. Unless otherwise noted, for the remainder of the results section the term "virtual acuity" refers to the line 4 scores.
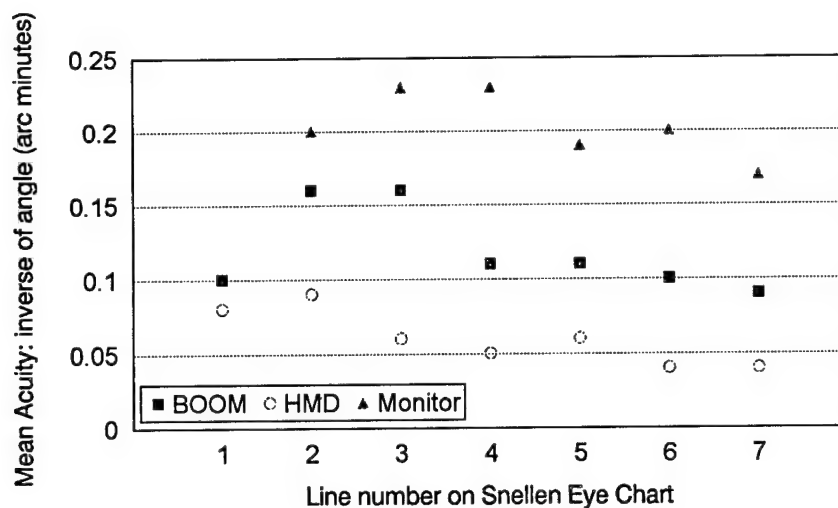
Figure 8. Virtual acuity for three display devices.

A one-way ANOVA was conducted to compare line 4 scores across display devices. A Least-significant difference (LSD) multiple range test indicated that the means for each of the display device groups differed significantly from the others ($F(2,44)=196.38$, $p < .01$).

Table 3 presents the descriptive statistics for the line 4 acuity scores. Note that the entry in the last column is not based upon the participants' acuity at 20 ft.

Table 3

Descriptive Statistics for Acuity Measure

Display	Mean	Median	S.D.	Range	Mean Angle Subtended	Mean Snellen Acuity
BOOM	.112	.100	.028	.060 - .180	10.	20/200
HMD	.048	.040	.017	.020 - .080	25	20/500
Monitor	.229	.220	.030	.180 - .280	4.55	20/91

The Pearson Correlation Coefficient for the real-world near and far visions tests was significant, $r = .47$, $n=48$, $p < .001$ for a two-tailed test. Partial correlation analyses controlling for display device were conducted between virtual acuity and real-world: near acuity, far acuity, and stereopsis. As shown in Table 4, none of these correlations approached statistical significance.

18

Table 4

Partial Correlation Analyses, Controlling for Group, Between Virtual Acuity and Real-world: Near Acuity, Far acuity, and Stereopsis.

	Coefficient	Degrees of Freedom	2-tailed Significance
near acuity	-.23	44	P= .120
far acuity	-.02	44	P= .873
stereopsis	.05	44	P= .727

Color vision. Table 5 lists for each group the mean scores for the two tests of color vision in the VE. Separate ANOVAs for each (red-green and yellow-blue) real-world color test found no differences between groups. For VE color visions tests, a one way ANOVA found a significant difference for yellow-blue color test scores among display devices $F(2, 42)=3.43$, $p<.05$. A Fisher LSD revealed that the BOOM produced significantly higher yellow-blue color test scores than the monitor. No significant difference was found among red-green scores obtained through the BOOM, HMD and Monitor presentation, $F(2,42)=1.073$, $p=.35$.

Table 5

Scores on Each Color Vision Test for Each Means of Presentation

Viewing condition	red-green	yellow-blue
Real World	3.00	2.91
BOOM	2.87	3.00
HMD	2.64	2.71
Monitor	2.62	2.37

The VE color test scores for each display device group were compared with real-world performance. Red-green scores were significantly lower in the HMD [$F(1,13)=7.22$, $p<.02$] and in the monitor [$F(1,15)=5.87$, $p<.03$] conditions than in the real world. No other differences were significant. A partial correlation, controlling for display device, indicated no significant correlation between virtual acuity and virtual color test score.

Object recognition. All participants correctly identified the object at the end of the corridor as a human figure. Although more detailed descriptions were not required for this task, some participants offered descriptions beyond the simple identification of the figure as human: 16

participants specified that the figure was a man, two participants identified the figure as a man with a guitar, another said it was a man with a gun, and one said the figure was a soldier.

Size (height) estimation. Table 6 lists descriptive statistics for height estimates of a human figure located 40 feet away, displayed either through the BOOM, HMD, monitor, or real-world conditions. Data for the real-world condition (Lampton, McDonald, Singer, & Bliss, 1995) were obtained from a separate group of 36 participants who performed the object recognition, size estimation, and distance estimation tasks in a real-world setting roughly comparable to the VE. (The area used for the real-world condition differed from the VE in that the ceiling was 11 feet high versus 12 feet in the VE and the corridor was 5 feet wide versus 3 feet in the VE.) One-sample t-tests revealed that the BOOM [t(15)=3.17, p<.01], HMD [t(15)=3.94, p<.001], monitor [t(15)=3.65, p<.002], and real-world [t(35)=4.42, p<.0001] viewing conditions produced estimates which were significantly lower than the actual height of the figure. For all viewing conditions, the participants underestimated the height of the figure.

Height estimates were compared between groups using a one-way ANOVA. The ANOVA revealed that the height estimates differed as a function of display condition, F(3,80)=5.32, p<.005. A Fisher LSD post hoc comparison revealed that the HMD produced height estimates significantly lower than the monitor and real-world conditions. No other differences were found between viewing conditions.

Table 6

Descriptive Statistics for Size (Height) Estimation (in.) of a 72-inch Figure

Display	Mean	Median	S.D.	Minimum	Maximum
BOOM	68.7	69.0	4.10	58.00	74.00
HMD	66.2	68.5	5.90	54.0	73.0
Monitor	69.0	69.5	3.28	60.0	73.0
Real world	70.5	71.0	1.95	67.0	74.0

Distance estimation. Distance estimation to the stationary figure at 40 ft was treated as a separate task from distance estimation of a moving figure starting from a known distance. Results for the stationary task are presented first.

Measures of central tendency for the distance estimates made at 40 ft in the real world and with the three VE display devices are shown in Table 7. An ANOVA comparing the VE and real-world display conditions was significant [F(3,80)=.4.9137, p=.0035]. An LSD range test

revealed that the mean distance estimate in the HMD condition was significantly greater than for each of the other conditions. No other conditions were significantly different.

Table 7

Distance Estimation (ft) at 40 ft by Display Conditions

Display (N)	Mean	Median	S.D.	Minimum	Maximum
Real World (36)	32.41	29.00	15.78	15.0	90.0
BOOM (16)	31.06	27.50	13.51	15.0	60.0
HMD (16)	77.63	30.00	91.97	6.0	300.0
Monitor (16)	33.91	30.00	19.13	10.0	90.0

As indicated by the median values, most participants underestimated the distance to the figure. However, with the HMD, several participants greatly overestimated the distance. Mean estimates were compared to a hypothetical perfect estimation value of 40 ft using a one-tailed t-test. Estimates made with the BOOM [$t(15)=-2.65$, $p<.02$] and in the real world [$t(15)=-2.88$, $p<.01$] were significantly less than the actual value of 40 feet. Neither the Monitor group nor the HMD group estimates differed significantly from the actual value. However, the variability of the HMD group estimates was very large, with a standard deviation more than four times larger than any other.

The data from the moving figure distance estimation task were analyzed with the following formula:

%Error = (E-A)/A

> where E is the target distance (the participant was instructed to call out when the figure reached the target distance)

> A is the actual distance at which the participant called out

Table 8 lists the item-total correlations for the five distances to be estimated. Cronbach's alpha was .72 for a reliability analysis of the VE estimates of the 5 distances. In an analogous reliability analysis of the real-world distance estimates Cronbach's alpha was .65.

Table 8

Item-Total Correlations for VE Estimates of 5 Distances

Distance to be estimated (ft)	Item-Total correlation
2.5	.58
5.	.84
10.	.82
20.	.61
30.	.56

Figure 9 presents percent error as a function of the distance to be estimated and display device. The percent error means for the VE conditions are negative values, indicating that in the VE conditions, unlike the real-world condition, the participants were calling out before the figure had reached the distance-to-be-estimated.
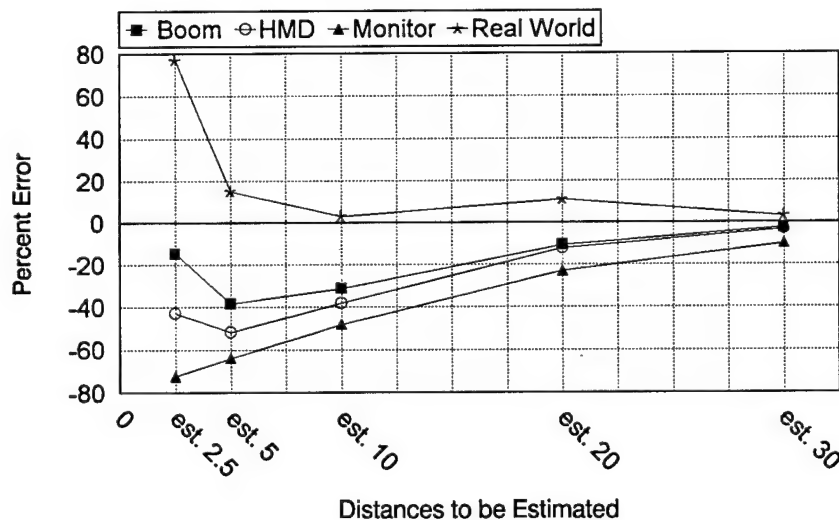

Figure 9. Percent error as a function of the distance to be estimated and display device.

An ANOVA was conducted with a between subject factor of display and a repeated measures factor of the five distances. The main effects of display device [$F_{(3,80)}=19.96$, $p<.001$], distance [$F_{(4,320)} = 7.58$, $p< .001$], and their interaction [$F_{(12,320)}=7.91$, $p<.001$] were significant.

An LSD range test collapsed across distances indicated that the real-world estimates differed significantly from the three VE conditions and that the BOOM was significantly better than the monitor. To examine distance estimation as a function of distance collapsed across devices, we used the absolute value of each percent error score to avoid averaging positive and negative values. An LSD range test indicated that distance perception was worse at 2.5 ft than at all other distances; at 5 ft perception was worse than 10, 20 and 30 ft, and perception at 10 ft was worse than 30 ft.

To examine the interaction of device and distance we conducted a separate LSD range test on the device means for each distance. For all distances, the real-world mean differed from each of the three VE display devices. At 30, 20, and 10 ft the BOOM was significantly better than the monitor. In addition, at 30 and 20 ft the HMD was significantly better than the monitor.

<u>Motor Tasks</u>

To determine if acuity using the VE display devices influenced performance on the VE motor tasks, partial correlation coefficients, controlling for display device, were computed for the virtual acuity score (at line 4) and the time and accuracy scores for the motor tasks. None of the correlations approached significance.

Reliability. Table 9 lists the Cronbach's Alpha measure of reliability for each of the motor tasks in which there were repeated trials.

Table 9

Reliability Analyses (Cronbach's Alpha) of VEPAB Tasks Across Trials

Task	N	Reliability (Alpha)	
		Time	Accuracy
Search	46	.65	-.07
Turns	47	.81	.56
Doorways	47	.87	.26
Bins	48	.85	.69
Choice RT	48	.65	.42
Tracking	48	N/A	.64

Table 10 lists the F values for a series of ANOVAs examining effects of display, practice, and display by practice (DxP) interactions for the search task and the locomotion tasks. Separate ANOVAs were conducted on the time and accuracy scores for the tasks that have both time and accuracy scores.

23

Table 10

F Values of ANOVAs for Time and Accuracy Scores for Each Motor Task

Task	Factor	Time	Accuracy
		F	F
Search	Display	9.31**	N/A
	Practice	2.79	N/A
	D x P	1.26	N/A
Turns	Display	.25	.50
	Practice	**11.28	4.88*
	D x P	.82	2.04
Doorways	Display	.28	.50
	Practice	2.00	.49
	D x P	1.88	1.34
Bins	Display	1.52	.57
	Practice	66.99**	7.65
	D x P	.60	.98
Choice RT	Display	.03	.95
	Practice	7.27**	3.71
	D x P	.43	.59
Tracking	Display	N/A	2.33
	Practice	N/A	9.96**
	D x P	N/A	.24
Figure-8	Display	.97	N/A

*p<.05   **p<.01

The Search task was the only motor task for which there was a significant effect for display device. Mean search times for the three devices are presented in Figure 10. An LSD range test indicated that the monitor group was significantly faster than either of the other display devices and the HMD was significantly faster than the BOOM.
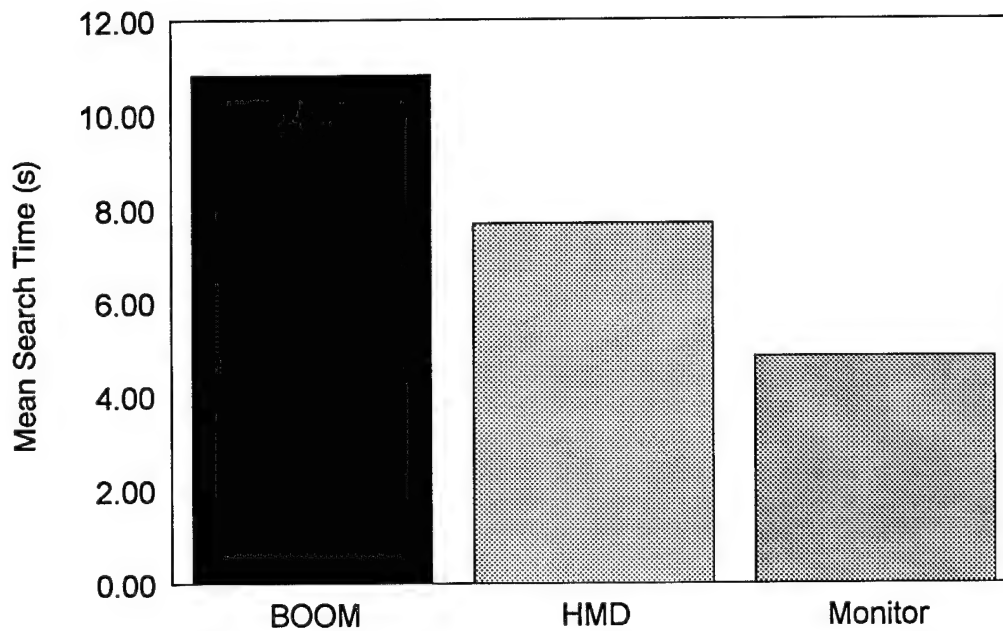
Figure 10. Mean Search task times for display devices.

Significant practice effects (performance on the last five trials was better than on the first five trials) were found for collision scores for the Turns task (Figure 11), time scores for the Turns, Bins, and Choice RT tasks (Figure 12), and percent time on target for the tracking task (Figure 13).
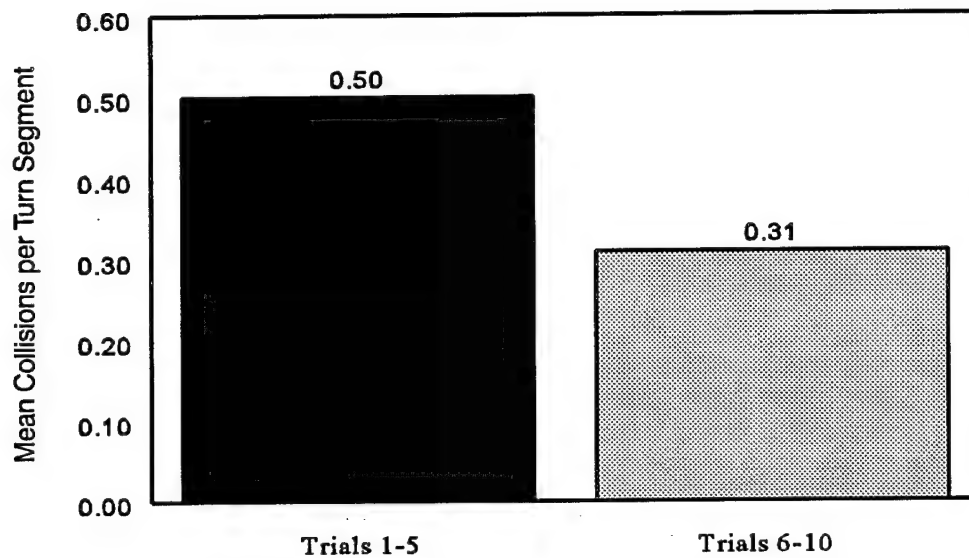
Figure 11. Mean collisions per segment for the first 5 trials vs. second 5 trials for the Turns task.
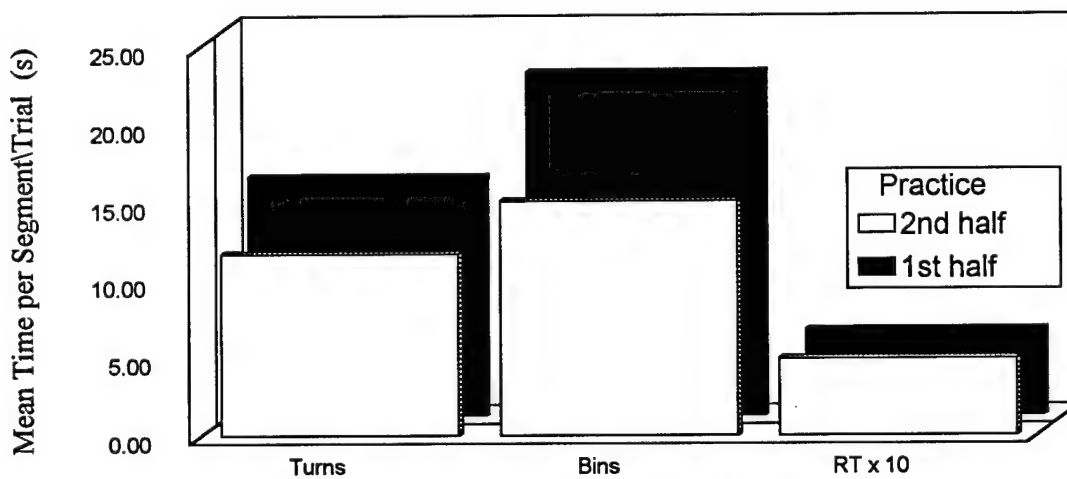
Figure 12. Mean time per segment\trial for the first 5 trials vs. second 5 trials for the Turns, Bins, and RT tasks.
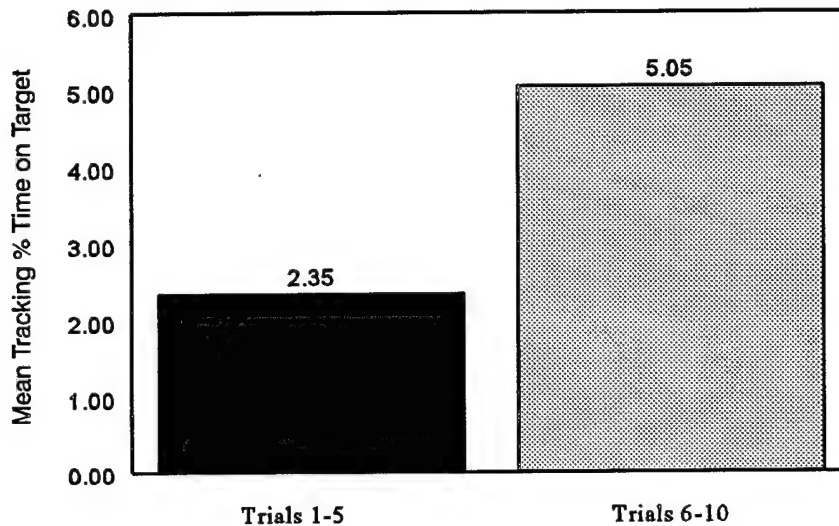
Figure 13. Mean %time on target per trial for the first 5 trials vs. second 5 trials for the Tracking task.

Measures of Aftereffects of VE immersion: Simulator Sickness Questionnaire and postural stability test

Individual items of the SSQ were converted to four scores, a Total Severity score and subscale scores for Nausea, Oculomotor discomfort, and Disorientation. For each of these scores, a repeated measures ANOVA was conducted with a between participants factor of display device and a within factor of the pre-immersion score and the post-immersion score. A similar pattern was found with each ANOVA: the pre- and post-immersion scores differed significantly and, effect of display device did not approach significance. For each set of scores the postscores were significantly higher than the pre. The pre and post means for Total Severity, and the Nausea, Oculomotor discomfort, and Disorientation subscales, are shown in Figures 14 to 17, respectively.
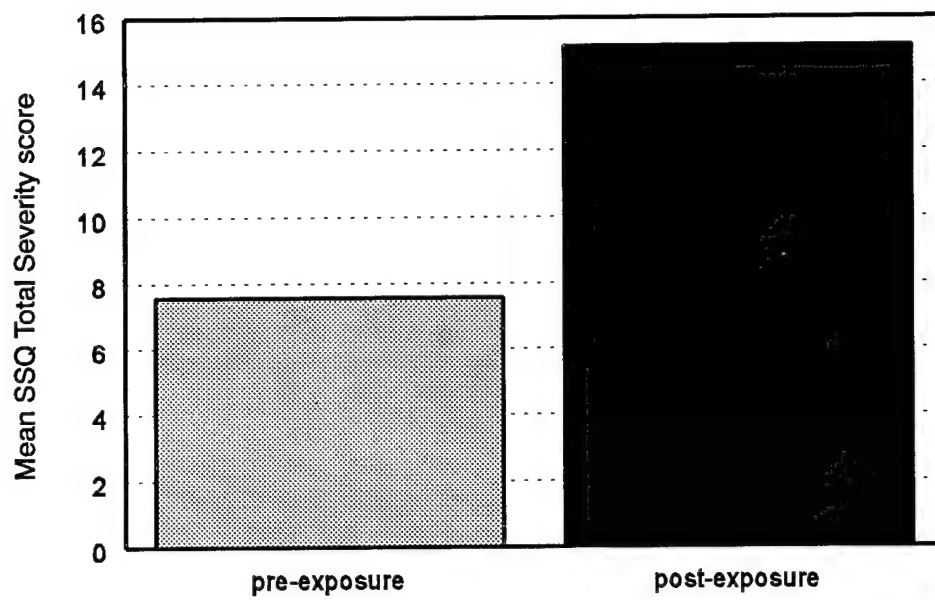
Figure 14. Mean SSQ Total Severity scores before (baseline) and after (postimmersion) performing VEPAB tasks.

Figure 15. Mean SSQ Nausea subscale scores before (baseline) and after (postimmersion) performing VEPAB tasks.
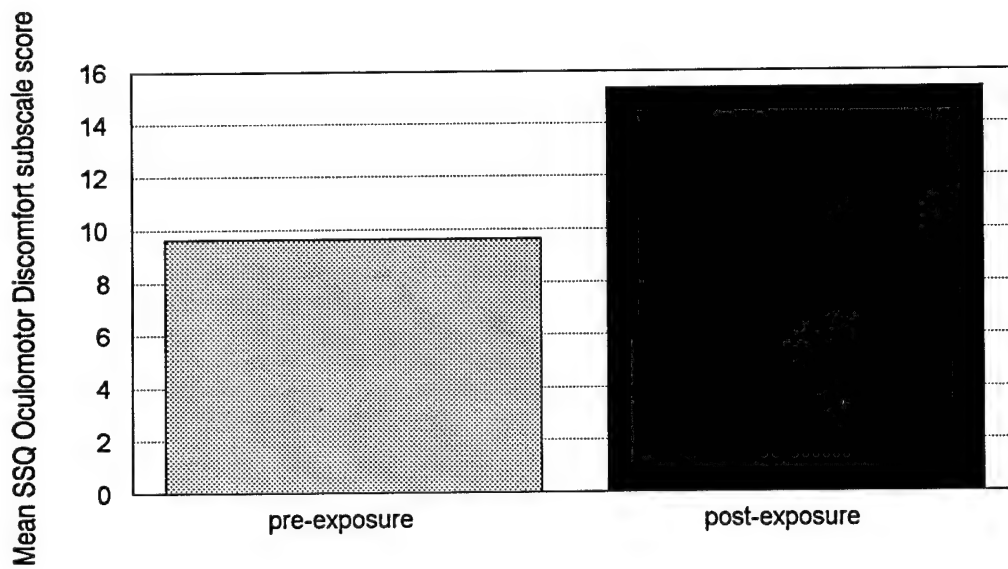
Figure 16. Mean SSQ Oculomotor discomfort scores before (baseline) and after (postimmersion) performing VEPAB tasks.
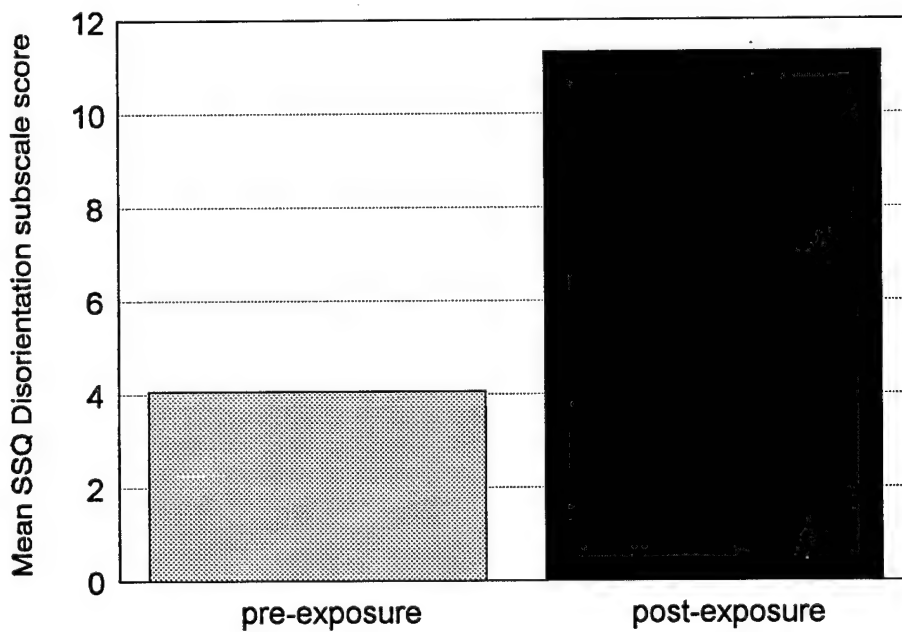
Figure 17. Mean SSQ Disorientation scores before (baseline) and after (postimmersion) performing VEPAB tasks.

A repeated measures ANOVA was conducted with a between participants factor of display device and a within factor of the p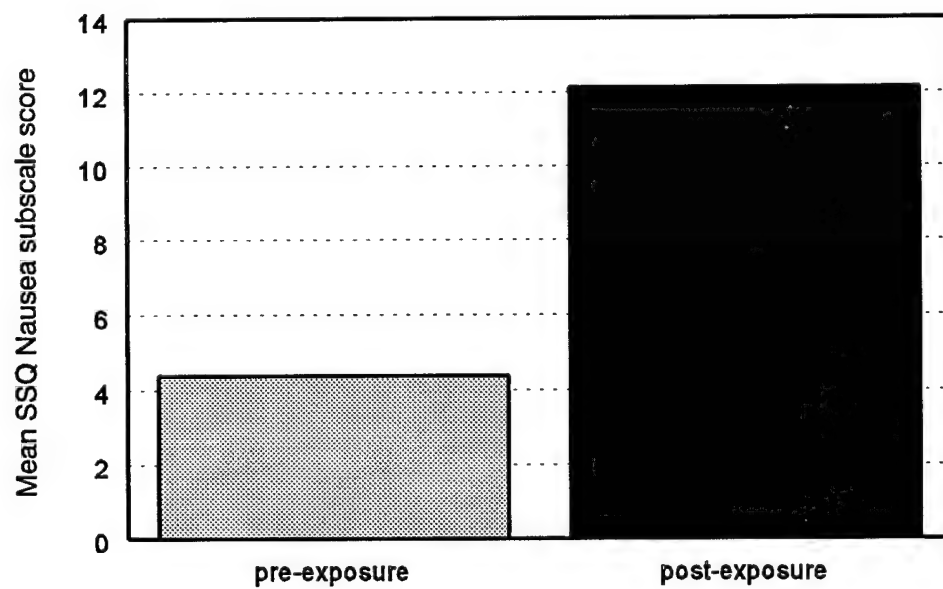re-immersion and the post-immersion postural stability scores. No significant effects were found for either factor or the interaction.

Severity ratings (none=0, 1=mild, 2=moderate, and 3=severe) for the individual symptoms on the SSQ were examined. Fatigue had the highest average severity rating for the BOOM (mean =.63) and monitor (mean =.50) groups. Eye strain (mean =.75) was the highest rated item for the HMD group.

<u>Correlations among task scores</u>

Partial correlations, controlling for device, were determined for the time and accuracy scores for Turns, Doorways, Bins, and RT. These correlations are presented in Table 11. Because the correlations were significant, or approached significance, only time scores were used in the next set of analyses.

Table 11

Partial Correlations, Controlling for Group, Between Speed and Accuracy Scores

Variables	Partial Correlation (degrees of freedom)
Turns: time and collisions	.65 (44) p=.000
Doorways: time and collisions	.31 (44) p=.038
Choice RT: time and correct choice	.28 (45) p=.052

Partial correlation coefficients, controlling for device, were computed for the time scores for the turns, doorways, tracking, bins, RT, search tasks, the Total Severity score derived from the Simulator Sickness Questionnaire (SSQ), and the background self-rating scores for weekly hours of use of computers and video games. The correlations coefficients are presented in Table 12.

Table 12

Correlations Among Tasks, Computer and Video Game Use, and Simulator Sickness

	Doorways	Tracking	Bins	RT	Search	video	computer	SSQ TS
Turns	.56 (43) P= .00	-.37 (44) P= .00	.64 (44) P= .00	.60 (44) P= .00	.16 (44) P= .29	-.23 (44) P= .12	-.09 (44) P= .54	-.01 (43) P= .94
Doorways		-.33 (44) P= .03	.47 (44) P= .00	.39 (44) P= .01	.18 (44) P= .24	-.17 (44) P= .26	-.30 (44) P= .04	.09 (43) P= .55
Tracking			-.62 (45) P= .00	-.42 (45) P= .00	-.26 (45) P= .07	.63 (45) P= .00	.28 (45) P= .06	-.26 (44) P= .08
Bins				.56 (45) P= .00	.37 (45) P= .01	-.32 (45) P= .03	-.33 (45) P= .02	.04 (44) P= .80
RT					.29 (45) P= .05	-.32 (45) P= .03	-.19 (45) P= .19	.16 (44) P= .28
Search						-.28 (45) P= .05	-.35 (45) P= .02	.06 (44) P= .70
video games							.35 (45) P= .015	-.24 (44) P= .101
computer								-.12 (44) P= .433

Note: (Coefficient / (D.F.) / 2-tailed Significance)

Discussion

Vision tasks

Acuity. The results of the measurement of the participants' acuity in VE can be summarized as follows: acuity measures were reliable; acuity varied across participants within each display group even though they had been screened for normal real-world vision; the rank order of the display devices based on the mean of the line 4 scores corresponded to the resolution (pixel per arc minute of angle) of the devices, but acuity was worse than would be predicted from pixel density alone.

Acuity scores were very reliable across the 7 lines of the VE Snellen eye chart. We suspect that some of the participants in the HMD group guessed, based upon previous experience with eye charts, that the top line of the chart was the letter "E" before the eye point had been moved close enough to the chart for the letter to be clearly discernable. Therefore, the mean distance at which the top line was read probably represents an overestimation of the acuity for the HMD. Because our criterion was that all the letters of a line must be recognized correctly, the mean score for line 7 may be an underestimation of acuity in that some participants could read many but not all of the letters at a given distance. (The same criterion was used for the real-world acuity tests). Some scheme could be devised to adjust the acuity scores for each line by calculating the percentage of the letters correctly identified. However, our use of the line 4 scores made this step unnecessary.

Acuity varied across individuals within display device groups. We suspect that some of the variation is attributable to individual differences in the ability to focus and maintain focus on images at short distances; the visual displays in the HMD and the BOOM are only a few inches away from the participants's eyes. VE acuity was not correlated with real-world acuity, however, the VE displays required focussing at shorter distances than are required in the real-world near point acuity test. Rinalducci (in preparation) pointed out that for HMD displays, focal length is constant but convergence changes when viewing the virtual space from near to far. Individuals may vary in their ability to maintain a fixed focal length while vergence is changing. For the monitor condition, participants were seated so that they would view the display from about 24 inches. However, the participants were not physically restrained from shifting position and the variability within the monitor group may be due in part to differences in the distance from which the screen was viewed.

The rank order of the mean acuity (based on the line 4 scores) for the devices corresponded to the order of the resolution (pixels per arc minute) of the devices. However, for each of the VE display devices the acuity was worse than would be predicted from the angle subtended by each pixel. (This is an admittedly simple approach, see Padmos and Milders (1992) for a discussion of approaches to estimating resolution from pixel density). Rinalducci, Cinq-Mars, Mapes, and Higgins (in preparation) empirically determined the angle subtended by each pixel in the HMD used in this experiment. Based on their findings, we estimate the resolution for the

HMD, expressed as the inverse of the minimum discriminable angle in arc minutes, as about .08 for the horizontal and about .096 for the vertical. In comparison, the average VE acuity was .048. Using the manufacturer's specifications for the FOV and the pixel density of the BOOM, a horizontal resolution of about .15 would be expected in contrast to the .112 average VE acuity that we observed. Using the manufacturer's specifications for the pixel density of the monitor, and assuming that the participants viewed the monitor from about 24 inches, the expected resolution is .58 in contrast to observed average VE acuity of .229.

In addition to the limits on resolution imposed by the pixel density of the displays, other factors could be expected to affect acuity. Rinalducci et al. (in preparation) pointed out that VE displays may lack appropriate brightness and contrast for optimal resolution. In addition, imperfections in the modeling of the VE and distortions in rendering could all interact to affect acuity. For example, the software used to render the VEPAB images performs anti-aliasing to manipulate the edges of an image to make them appear smoother to the eye. Anti-aliasing reduces the stairstep effect which would otherwise appear with diagonal lines, thus smoothing, but also blurring, the image.

For the devices with which we present stereoscopic images, the BOOM and HMD, there are additional factors which may affect acuity. Robinett and Rolland (1992) described the complex challenge of computing correct stereoscopic images for HMDs. For the monitor, unlike the HMD and BOOM, glare from the lights in the room could not be completely eliminated. In addition, for determining acuity for the monitor, it is not clear to us as to how the distance from the participant to the screen in the real-world should be considered relative to the distance from the simulated eye point to the eye chart in the VE.

Characteristics of display devices such as pixel density, brightness, and contrast interact with characteristics of participants, such as the ability to focus at short distances, to affect VE acuity. As a result of all of these factors, the specifications provided by the manufacturers of VE display devices did not provide accurate estimates of the participants' acuity with the display.

Color perception. There were two primary findings with the color tests: (a) for each of the three displays some of the participants who had perfect scores on the real-world color vision tests did not correctly identify all the numerals in the VE color vision tests; and (b) one of the two participants who could not pass the real-world color vision test could identify the numerals in VE. The latter finding is consistent with pilot testing in which real-world color weak participants could easily read the numerals when displayed on a monitor.

To fail color blind participants appropriately, and to pass participants with normal real-world color vision, the VE color test must present -- with both appropriate color and brightness -- the dots that define the numerals and the other dots in the random pattern in which the numerals are imbedded. The obvious failure to do so may be based on imperfections in the method in which the color test plates were digitized through scanning, errors in software

33

rendering, imperfections in the display devices, or an interaction of these factors. Boff and Lincoln (1988) noted that the luminance and flicker of a display can affect color appearance. Inconsistent representations of color in VEs have implications for military training applications; for example, variation in color presentation across devices or within a device across sessions may result in training exercises in which the detectability of camouflaged troops or vehicles varies greatly and inappropriately for training purposes.

Object recognition. For the three VE display device conditions and the real-world viewing condition all participants identified that object at the end of the 40-ft hallway as a human figure. Although the task was not sensitive to differences in the display conditions it was still worthwhile to administer because one purpose of this task is to provide the context of a familiar figure for the subsequent size and distance estimation tasks. Obviously, we could obtain a more sensitive task by requiring the participants to describe the figure in detail, which some of the participants do spontaneously anyway. Quantifiable measures would include: recognition that the figure is a soldier, that the figure is holding an object, and that the object is a gun. The measures could be dichotomous (yes/no the participant recognized the object as a gun) and would be interesting to compare with the acuity measures.

Size (height) estimation. The mean height estimates in VE were consistently shorter than the actual height of the figure. This may be a function of the VE ceiling height of 12 ft. Participants may have assumed that the ceiling was about 10 ft high. The six-ft human figure, one-half of the height of the corridor, may therefore have appeared shorter in the context of the ceiling height.

Distance estimation. For the static distance estimation task, distance perception in VE was marked by great variability across participants, particularly for the HMD. As in the other display conditions most HMD participants underestimated the distance to the figure. This finding is consistent with research on distance estimation in flight simulators (Wright, 1995). However, several participants in the HMD condition greatly overestimated the distance. It is possible that the HMD condition does not convey enough cues to distance to allow more than random guesses. However, a possible explanation of why some participants overestimated distance in VE relates to the characteristics of the HMD display. Murch (1973) (in the context of real-world perception) listed contrast, clarity, and brightness among monocular cues to depth. The contrast, clarity, and brightness of the figure presented via the HMD display may be relatively weak and perceived by some participants as indicating that the figure is very far away. The range of brightness and contrast is limited in VE display systems (McKenna and Zeltzer, 1992).

Although distance estimation in the real world was less variable than in VE, for the static distance estimation task it differed significantly from perfect performance and, like the VE conditions, involved underestimation of distance. However, for distance perception involving a moving figure, in which the participants are told the initial distance to the figure, real-world performance differed greatly from VE. In the real world, most participants called out after the

34

figure had closed to or passed the specified distance. In general, distance perception was relatively good in the real world except for the estimates of 2.5 ft, for which several participants allowed the figure to approach within inches before calling out.

In the VE conditions most participants called out before the figure had closed to the specified distances. Because the pattern of errors in the VE differs from the real-world condition, we do not believe that the procedure itself is solely responsible for the similar pattern we observed across all three of the VE displays.

Motor tasks

Reliability. Moderate to high reliability was found for each of the motor task's time performance scores across trials. Reliability of accuracy scores varied from moderate to low. The lowest reliability was found for the accuracy measure of the Search task. The accuracy measure for the Search task was arbitrarily defined as whether or not the participant detected the target within the 45-second time limit. On almost all trials the target was spotted well within the time limit. Therefore, the variance of the search accuracy scores was near zero and the reliability score was low.

Correlations among tasks. For the Turns and Doorways tasks speed and accuracy had significant positive correlations, indicating that there was not a tradeoff of speed for accuracy. In the locomotion tasks, collisions almost always result in a delay, thus the VEPAB tasks do not encourage "bounce off the walls" strategies found with some video games. For the choice reaction time task, the results suggested (the correlation approached significance) a tradeoff between the speed and accuracy of responding, similar to that which might be expected in a real-world RT task. The RT task, unlike the locomotion tasks, did not require precise control of movement.

Video game and computer use self-ratings were highly correlated. Video game use was significantly correlated with Tracking, Bins, and RT tasks and the correlation approached significance with the Search task. Computer use was significantly correlated with Doors, Bins, and Search, and approached significance with the Tracking task. For all of these correlations, more video game or computer use was associated with better (faster or more accurate) task performance. SSQ Total Severity scores were not significantly correlated with any of the motor tasks nor with computer or video game use.

Practice effects. Small but statistically significant practice effects, comparing performance on the first five trials versus the last five trials, were found for the Turns, Bins, RT, and Tracking tasks. These findings were consistent with our previous research with VEPAB (Lampton, Knerr, et al., 1994). We suspect that the practice effect is based, in part, upon increasing skill in manipulating the joystick, adaption to the VE display, and refining strategies for successfully completing tasks.

35

Performance differences across devices. The Search task was the only motor task for which we clearly expected the devices with head tracking, the BOOM and the HMD, to be significantly better than the monitor. Our expectation was based on the following line of thought: The challenge of the Search task is to quickly slew the direction of regard so that the target is within the FOV. Once the target is within the FOV, it is easily detectable even with the display device with the lowest resolution, the HMD. Controlling the direction of regard through head movement would be more natural, more highly practiced, and therefore lead to better performance than control with a joystick.

However, for the Search task the monitor group was significantly faster than the BOOM and HMD. Several factors encouraged participants using the BOOM or the HMD to employ slow, steady head movement. For safety reasons, participants were instructed to avoid very rapid head movement. Also, the HMD and BOOM have enough inertia to slow head movement. Finally, we suspect that rapid slewing of the FOV with the HMD and, to a lesser extent the BOOM, may result in a noticeable lag between the position of the head and the position of the FOV. This sort of lag is unpleasant, can produce simulator sickness, and may discourage the participant from making rapid movements.

With the monitor condition the participant controlled the direction of regard by pushing the joystick to the left or right to look from side to side, and forward or back to look up or down. System software limited the maximum slew rate. The participants could very quickly scan large segments of the room. The participant had only to see the target flash through their FOV to successfully complete a trial. The participant was not required to report where the target was (in azimuth or elevation) in relation to the participant's initial position in the VE. Thus, the performance measure for the Search task did not represent the kind of performance that may be required in VE-based training: determining the position of a target relative to the participant.

For the Tracking task, the devices with stereoscopic displays, the BOOM and the HMD, might be expected to allow better performance than the Monitor because the target moved in three dimensions. The perception of depth should aid in tracking when the target moves toward or away from the eye point. Unfortunately, the slowness of the maximum speed at which the tracking cursor could move overshadowed any effect of the visual display.

For the motor tasks other than the Search and Tracking tasks we did not expect and we did not find significant differences across display devices. High resolution or wide FOV is not required for the successful completion of the tasks. The tasks involve visual stimuli that are large and readily visible. For the locomotion tasks there are multiple cues to support task performance. If the locomotion tasks were changed to require running rather than walking through the VE space then the resolution and FOV of the devices might be expected to affect task performance. Running without collisions would require more rapid and precise perception of the spatial relation of obstacles and require greater capability to preview the area to be traversed.

36

Simulator Sickness

There were three main findings regarding simulator sickness. These were: simulator sickness was less severe than in our previous VE experiments; the monitor condition produced some discomfort; and the results suggested that there may be differences in the individual symptoms across devices (although the SSQ subscale scores did not differ across devices).

In previous VE experiments reviewed by Lampton, Kolasinski, et al. (1994), between 4% and 16% of the participants experienced discomfort to the extent that their participation in the experiment was terminated. For this experiment, no participant reported or manifested symptoms to the extent that we had to consider stopping their participation in the experiment. This experiment differed from the others along several dimensions that might be expected to influence simulator sickness, for example, the cumulative time of immersion and the schedule of breaks. However, we suspect that this improvement is based in part upon the use of a more powerful computer which increased the update rate of the VE images.

We did not expect the monitor condition to produce symptoms of simulator sickness. Working at a computer monitor for hours on end might be expected to produce eye strain or fatigue. However, in this experiment the total time for performing the VEPAB tasks at the monitor was less than an hour and that included rest breaks. In the monitor condition, the VE display has a narrow FOV and the participants can shift their gaze from the VE to the real world at any time. Both of these factors could be expected to produce less simulator sickness than the BOOM or HMD conditions. An explanation is that the tasks that involve vection, the perception of self-movement, may produce some symptoms of simulator sickness even when presented on the monitor. This results from a mismatch between visual information which conveys movements, and vestibular and kinesthetic information which does not. For future research, we may test this hypothesis by comparing simulator sickness after equal durations of immersion performing either the VEPAB Figure-8 task and the Bins task. If immersion itself is producing symptoms then there should be no differences. If vection is a necessary component then the Figure-8 group should report more simulator sickness.

Although the subscale scores did not differ significantly across display devices, the results suggest that the incidence of particular symptoms may differ across devices. For example, the symptom "nausea" seemed more prevalent with the BOOM than in the other devices. The BOOM differs in several ways from the other devices, including having the widest FOV of the three visual display devices used in this experiment. Overall, we think the results demonstrate that the SSQ is a sensitive measure of discomfort and that our civilian participants are not reluctant to report even minor perceptions of discomfort.

37

Changes to VEPAB

The VEPAB tasks have proven to be useful for assessing the effects of interface characteristics on human performance in virtual environments (Lampton, Knerr, et al. 1995; Singer, Ehrlich, Cinq-Mars, & Papin, 1995). These or similar tasks should also prove useful in the future when design tradeoffs must be made in the process of developing training system prototypes. However, in the course of our research we have identified several changes to the tasks to improve their utility.

For future research involving the Search task, the performance measure should be changed to require the participant to specify the azimuth and elevation of the target relative to the participant. Researchers at the Naval Air Warfare Center Training Systems Division have adapted the VEPAB search task for use in a systematic program to examine orientation in VEs. We hope that information from that program can be used to guide additional improvements in the VEPAB search task.

The tracking task was very difficult. In many trials, the participants never managed to get the cursor to the target before the time elapsed. A major problem with the tracking task is that the VEPAB software sets a maximum rate for cursor movement which is too slow to allow adequate tracking performance. The maximum speed at which the cursor can move is not much faster than the speed at which the target moves. It may be that in VE tracking a target that moves and can change directions in three dimensions is a difficult task anyway, but the slow cursor speed did not allow a good test of tracking in VE.

The VEPAB distance estimation tasks were sensitive to differences in individual performance and different display devices. However, both the VE used in the task and the task procedure could be improved. Although the VE was modeled after an actual office building, the 12-ft ceiling is probably atypically high and, combined with a narrow hallway, may make the distance estimation task more difficult than would be found in most VE training applications. In addition, the homogeneous gray background at the end of the hallway should be replaced by a wall similar in appearance to the sides of the hallway. Regarding the distance task procedure, emphasis on producing tasks that could be quickly administered resulted in a dynamic distance estimation task that has characteristics of both estimation and production procedures for measuring distance perception. As such, the task is more complex than was intended. Finally, unlike the VEPAB motor tasks, the vision tasks, including distance estimation, were administered before participants were allowed to practice interacting with the VE. Perhaps distance estimation should be measured two or more times; upon initial immersion, before the participants have had much time to experience the VE, and later after participants have been given time in the VE.

Our colleagues Bob Witmer and Paul Kline at the Simulator System Research Unit have initiated a research program specifically to address factors, such as field of view and texture,

thought to affect distance perception in VEs related to training dismounted soldiers. Results from that research may indicate ways to improve the VEPAB.

Although the VEPAB tasks could conceivably be used with any visual displays or control devices, the virtual environments of VEPAB are already showing their age. That is, although the VEPAB VEs were intentionally designed to be simple, they now lack the level of detail that can be routinely achieved with current VE modeling and rendering hardware and software. Considerable improvements in many of the software aspects of generating and displaying VEs have occurred since the VEPAB VEs were programmed in 1992. The simple VEPAB VEs will probably continue to be an efficient way to quickly train participants to operate the VE control devices with which they interact with the VE. However, if VEPAB is to continue to be a viable tool for supporting research into training applications of immersive VEs, then eventually extensive recoding may be necessary. In addition, several new tasks could be added. Locomotion tasks involving running and crawling would be beneficial for research involving dismounted infantry. Tasks involving the identification and localization of sounds would likewise support training research. Finally, a task to empirically measure FOV in the VE is needed, given the confusing interactions of hardware and software in determining the FOVs of VEs.

Conclusion

The VEPAB vision tasks were sensitive to differences in display devices and to individual differences. Within device conditions, the variation in VE acuity across individuals demonstrates the need to empirically measure acuity and not just rely on the resolution specifications provided by the device manufacturer. Visual display devices produced acuity scores ordered according to the horizontal pixel density of the displays, but the acuity scores were worse than what would be predicted solely on the resolution alone. We believe this is a result of inaccurate manufacturer specifications of pixel density and FOV, errors in image modeling and rendering, low brightness and contrast in the displays, optical distortions, and perhaps incompatible demands of accommodation and convergence.

Size and distance estimations were more variable in VE than analogous real-world tasks. In the VE, size and distance were underestimated for a nonmoving figure at 40 ft. In a task in which participants were to call out when a moving target closed to specified ranges, with the VE displays most participants called out before the target had reached the specified distances. In the real-world, most participants called out as or after the target reached the specified distances.

Counter to our expectations, mean search time with the monitor was faster than for the displays with head tracking. For head mounted displays, inertia and safety instructions may have led to slower times. In addition, the performance criterion for the task required only the detection of the target. Adding a requirement to determine the target's position relative to the

participant would make the task more representative of the kind of performance that we expect will be required in VE training applications.

The motor tasks had high reliability and demonstrated small but significant practice effects. These findings were consistent with previous research (Lampton, Knerr, et. al, 1995). Unlike the vision task scores, most of the motor tasks were correlated with participants' use of computers and video games, with higher use associated with better performance.

Although total immersion time was less than an hour, including several breaks, we still found a significant increase in simulator sickness in comparing pre-immersion and post-immersion scores. Unexpectedly, even the monitor group showed a significant increase in simulator sickness scores. We believe that the VEPAB locomotion tasks induce vection, the perception of self-motion, and that the discrepancy between the visual system, which signals movement on the part of the participant, and other sensory systems which do not, result in simulator sickness.

VEPAB has been demonstrated to provide a means of familiarizing participants with perceptual and control interfaces for interacting with VEs. VEPAB also provides measures of those perceptual and motor skills. By providing benchmarks of human performance, VEPAB can promote continuity in training research involving different technologies, separate research facilities, and dissimilar subject populations.

Burdea and Coiffet (1994) pointed out that research in older fields which were the precursors to Virtual Reality, such as teleoperation, lacked general evaluation procedures. As a result, much of the research in those areas failed to obtain results applicable to systems other than the particular system being studied. In addition, a standard practice was to focus on the average scores from user tests, ignoring individual differences. They predicted that VR evaluation studies may have these same weaknesses. We believe that VEPAB provides tools that can help research on training applications of immersive VEs avoid problems such as those identified by Burdea and Coiffet.

References

Alluisi, E. A. (1991). The development of technology for collective training: SIMNET, a case history. *Human Factors, 33*(3), 343-362.

Boff, K. R., & Lincoln, J. E. (1988). *Engineering data compendium: Human perception and performance.* AAMRL, Wright-Patterson AFB, OH.

Burdea, G., & Coiffet, P. (1994). *Virtual reality technology.* New York: Wiley.

Geldard, F. A. (1972). *Visual acuity and contrast. The human senses* (2nd ed.), 125-153. New York: Wiley.

Gorman, P. (1991). *SuperTroop via I-Port: Distributed simulation technology for combat development and training development* (IDA Paper P-2374). Alexandria, VA: Institute for Defense Analysis.

Kennedy, R. S., Lane, N. E., Lilienthal, K. S., Berbaum, K. S., & Hettinger, L. J. (1992). Profile analysis of simulator sickness symptoms: Application to virtual environment systems. *PRESENCE: Teleoperators and Virtual Environments, 1*(3), 295-301.

Knerr, B. W., Goldberg, S. L., Lampton, D. R., Witmer, B. G., Bliss, J. P., Moshell, J. M., & Blau, B. S. (1993). Research in the use of virtual environment technology to train dismounted soldiers. *Proceedings of the 15th Interservice/Industry Training Systems and Education Conference,* Orlando, FL.

Lampton, D. R., Knerr, B. W., Goldberg, S. L., Bliss, J. P., Moshell, J. M., & Blau, B. S. (1994). The virtual environment performance assessment battery (VEPAB). Development and evaluation. *PRESENCE: Teleoperators and Virtual Environments, 3*(2), 145-157.

Lampton, D. R., Knerr, B. W., Goldberg, S. L., Bliss, J. P., Moshell, J. M., & Blau, B. S. (1995). *The virtual environment performance assessment battery (VEPAB): Development and evaluation* (ARI Technical Report 1029). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A297 277)

Lampton, D. R., Kolasinski, E. M., Knerr, B. W., Bliss, J. P., Bailey, J. H., & Witmer, B. G. (1994). Side effects and aftereffects of immersion in virtual environments. *Proceedings of the Human Factors Society, 38th Annual Meeting,* 1154-1157.

Lampton, D. R., McDonald, D. P., Singer, M. J., & Bliss, J. P. (1995). Distance estimation in virtual environments. *Proceedings of the Human Factors and Ergonomics Society, 39th Annual Meeting,* 1268-1272.

41

Levison, W. H., & Pew, R. W. (1993). *Use of virtual environment training technology for individual combat simulation* (ARI Technical Report 971). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A263 546)

McCormick, E. J., & Sanders, M. S. (1976). *Human factors engineering and design.* New York: McGraw-Hill.

McKenna, M., & Zeltzer, D. (1992). Three dimensional visual display systems for virtual environments. *PRESENCE: Teleoperators and Virtual Environments, 1*(4), 421-458.

Moshell, J. M., Blau, B. S., Knerr, B. W., Lampton, D. R., & Bliss, J. P. (1993). A research testbed for virtual environment training applications. *Proceedings of the IEEE Virtual Reality Annual International Symposium,* 83-89.

Murch, G. M. (1973). *Visual and auditory perception.* New York: Bobbs-Merril.

Norusis, M. (1995). *SPSS 6.1 guide to data analysis.* Englewood Cliffs: Prentice-Hall.

Olzak, L. A., & Thomas, J. P. (1986). Seeing spatial patterns. In K. R. Boff, L. Kaufman, & J. P. Thomas (eds.), *Handbook of perception and human performance, vol. 1: Sensory processes and perception.* Chapter 7.

Padmos, P., & Milders, M. (1992). Quality criteria for simulator images: A literature review. *Human Factors, 34*(6), 727-748.

Pimental, K., & Teixeira, K. (1993). *Virtual reality: Through the new looking glass.* New York: McGraw-Hill.

Rinalducci, E. J. (in preparation). *Characteristics of visual fidelity in the virtual environment.*

Rinalducci, E. J., Cinq-Mars, S. G., Mapes, D., & Higgins, K. E. (in preparation). *Determining the field of view in HMDs: A psychophysical method.*

Robinett, W., & Rolland, J. P. (1992). A computational model for the stereoscopic optics of a head-mounted display. *PRESENCE: Teleoperators and Virtual Environments, 1*(1), 45-62.

Singer, M. J., Ehrlich, J., Cinq-Mars, S. G., & Papin, J. (1995). *Task performance in virtual environments: Stereoscopic vs. monoscopic displays and head-coupling* (ARI Technical Report 1034). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sterling, B. (1993). War is virtual Hell. *WIRED*, premiere issue, 46-51, 94-99.

Witmer, B. W., & Singer, M. J. (1994). *Measuring presence in virtual environments* (ARI Technical Report 1014). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A286 183)

Wright, R. H. (1995). *Virtual reality psychophysics: Forward and lateral distance, height, and speed perceptions with a wide-angle helmet display* (ARI Technical Report 1025). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A294 027)